Vlerick Leuven Gent Management School

the Autonomous Management School of
Ghent University and Katholieke Universiteit Leuven

Vlerick Leuven Gent Working Paper Series 2006/03

EVIDENCE AND IMPLICATIONS OF ZIPF'S LAW

FOR INTEGRATED ECONOMIES

HARRY P. BOWEN

Harry.Bowen@vlerick.be

HARIS MUNANDAR

JEAN-MARIE VIAENE

D/2006/6482/3

EVIDENCE AND IMPLICATIONS OF ZIPF'S LAW

FOR INTEGRATED ECONOMIES

HARRY P. BOWEN

Harry.Bowen@vlerick.be

HARIS MUNANDAR

JEAN-MARIE VIAENE

We thank Xavier Gabaix for helpful comments and clarifying discussions. We also benefited from comments from Charles van Marrewijk, Enrico Pennings, participants of the European Trade Study Group (Dublin) and seminar participants at Erasmus University and the Tinbergen Institute.

Contact:

Harry P. Bowen

Vlerick Leuven Gent Management School

Tel: +32 16 32 35 86

Fax: +32 16 32 35 80

Email: Harry.Bowen@vlerick.be

ABSTRACT

This paper considers the distribution of output and productive factors among members of a fully integrated economy (FIE) in which there is free mobility of goods and factors among members and whose members share the same technology. We first demonstrate that, within an FIE, each member's share of total FIE output and its shares of total FIE stocks of productive factors will be equal. If economic policies are largely harmonized across FIE members then this "equal-share" property implies that the growth in any member's shares of FIE output and factor stocks can be taken to be a random outcome. Building on Gabaix's (1999) result for the distribution of city sizes we argue that, if output and factor shares among FIE members evolve as geometric Brownian motion with a lower bound, then the limiting distribution of these shares will exhibit Zipf's law. We empirically examine for Zipf's law for the distribution of output and factor shares across two (presumably) integrated economies: the 51 U.S. states and 14 European Union (E.U.) countries. Our empirical findings strongly support Zipf's law with respect to the distribution of output, physical capital and human capital among U.S. states and among E.U. countries. These findings imply that models used to characterize the growth of members within an FIE should embody a key assumption: that the underlying growth process of shares is random and homogeneous across FIE members.

JEL Classification: E13, F15, F21, F22, O57

Keywords: growth, economic integration, Zipf's law.

EVIDENCE AND IMPLICATIONS OF ZIPF'S LAW FOR INTEGRATED ECONOMIES

The number of regional trade agreements has increased continuously since the early 1990s and many new initiatives for special association agreements are currently being negotiated (see WTO website). Institutional arrangements under which countries open their borders differ in reality. Most agreements are designed to increase international trade between markets but a few, like the European Union, also allow greater mobility of productive factors within the integrated area. In the limit, such integration would be represented by a fully integrated economy (FIE) in which there is free mobility of goods and factors among FIE members together with complete harmonization of economic and social policies.

While prior work has demonstrated the potentially important role of trade[1] and factor mobility[2] as influences on economic growth, less attention has been given to the question of how trade and factor mobility impact the distribution of output across members of a FIE, and hence how these influences affect the relative economic position of members. Apart from being simply a question of distributional consequences, analysis of this question has important implications for the nature of models that can be used to characterize the growth processes of FIE members. Specifically, as we will show in this paper, the distribution of output and factor shares across FIE members can be expected to conform to a rank-share distribution that exhibit *Zipf's law*, which indicates a specific relationship between the rank and value of a variable. This result implies that models used to characterize the growth of members within an FIE must embody a key

[1] An extensive body of work has explored the role of international trade and of factor mobility as mechanisms generating endogenous economic growth. For example, Grossman and Helpman (1991) show that trade generally enhances growth, particularly when it facilitates the international transmission of knowledge. Similarly, Rivera-Batiz and Romer (1991) show that increased trade due to economic integration may have both level and growth effects depending upon the processes by which R&D and information flow across borders. Devereux and Lapham (1994) extend Rivera-Batiz and Romer's model to show that, even without knowledge flows, the balanced growth rate when there is free trade in goods alone exceeds that in autarky, provided that initial levels of national income differ across countries.

[2] For example, Baldwin and Martin (2004) examine the relationship between growth and the agglomeration of economic activity and find that it depends crucially on the extent of capital mobility between regions. Similarly, Viaene and Zilcha (2002) show that while complete capital market integration among countries has a positive effect on outputs, it does not raise long-run growth rates above autarky values. Instead, these growth rates are affected only by parameters that describe the accumulation of human capital.

assumption: that the underlying growth process of shares is random and homogeneous across members.

Country shares of regional (or world) output, or shares of a region's total supplies of productive factors, have become increasingly important constructs in the international trade literature (e.g., Bowen et al., 1987; Helpman and Krugman, 1985; Leamer, 1984; Viaene and Zilcha, 2002). In this regard, below we first demonstrate that, within an FIE, each member's share of total FIE output will equal its shares of total FIE stocks of each productive factor (i.e., physical capital and human capital). If economic policies are largely harmonized across members then this equal-share property implies that the growth in any member's shares of FIE output and factor stocks can be taken to be a random outcome. Following Gabaix (1999), if output and factor shares evolve as geometric Brownian motion with a lower bound, then the limiting distribution of these shares will exhibit Zipf's law. Given this result, we then show that the limiting values of each FIE member's output and factor shares are completely determined once the number of FIE members is specified.

Given the theoretical expectation of Zipf's law for output and factor shares, we empirically examine for this law within two (presumably) integrated economies: the 51 U.S. states and 14 countries of the European Union (E.U.). The data generally cover the period from 1965 to 2000. Our empirical results convincingly support Zipf's law for U.S. states and for E.U. countries.

EQUALITY OF OUTPUT AND FACTOR SHARES IN INTEGRATED ECONOMIES

To demonstrate the equality of output and factor shares for each member of a fully integrated economy we consider an integrated economy that consists of $m = 1, \ldots, M$ members, each producing a single good by means of a constant return to scale production function of the form:

(1) $$Y_{mt} = F(K_{mt}, H_{mt}) \qquad m = 1, \ldots, M$$

where Y_{mt} is the level of output, K_{mt} the stock of physical capital, and H_{mt} the stock of human capital, all for country m at time t. The production function is assumed to satisfy all the neoclassical assumptions including diminishing marginal productivity with respect to each factor. For ease of exposition, the production function is assumed to take the Cobb Douglas form:[3]

$$(2) \qquad Y_{mt} = A_{mt} K_{mt}^{\alpha_m} H_{mt}^{1-\alpha_m} \qquad m = 1, \ldots, M$$

where A_{mt} is a scale parameter and α_m is capital's share of total output. If physical capital and labor are perfectly mobile between the M economies then we would expect the marginal product of each factor to be equal. Barriers to capital mobility (e.g. corporate income tax differentials, capital controls) or labor mobility (e.g. language, different pension systems) would instead create persistent differences in factor rates of returns between members. Consider one reference member of this integrated economy that, without loss of generality, we take to be country i. Let λ_{mt}^{k} and λ_{mt}^{h} define the proportional difference in rates of return to physical capital and to human capital between any country m and reference country i. The relation between rates of return to physical capital in the integrated economy can then be written as:

$$(3) \qquad v_1 \lambda_{1t}^{k} \frac{Y_{1t}}{K_{1t}} = \ldots = \frac{Y_{it}}{K_{it}} = \ldots = v_M \lambda_{Mt}^{k} \frac{Y_{Mt}}{K_{Mt}}$$

where $v_m = \alpha_m/\alpha_i$, implying $v_m = 1$ when $\alpha_m = \alpha_i$ ($m = 1, \ldots, M$). Note that for $m = i$, $\lambda_{it}^{k} = 1$ and $v_i = 1$. Likewise, the relation between rates of return to human capital can be written:

$$(4) \qquad \omega_1 \lambda_{1t}^{h} \frac{Y_{1t}}{H_{1t}} = \ldots = \frac{Y_{it}}{H_{it}} = \ldots = \omega_M \lambda_{Mt}^{h} \frac{Y_{Mt}}{H_{Mt}}$$

[3] The Cobb-Douglas specification has wide empirical support (e.g., Mankiw et al., 1992). The analysis can be extended to the case where the production function has the constant elasticity of substitution (C.E.S.) form.

where $\omega_m = (1-\alpha_m)/(1-\alpha_i)$, implying $\omega_m = 1$ when $\alpha_m = \alpha_i$ (m = 1, ..., M). Note that for $m = i$, $\omega_i = 1$ and $\lambda_{it}^h = 1$. The ratio of (3) to (4) gives the following relationship between ratios of human to physical capital:

$$(5) \qquad \eta_1 \lambda_{1t} \frac{H_{1t}}{K_{1t}} = ... = \frac{H_{it}}{K_{it}} = ... = \eta_M \lambda_{Mt} \frac{H_{Mt}}{K_{Mt}} = \frac{\sum_{m=1}^{M} \eta_m \lambda_{mt} H_{mt}}{\sum_{m=1}^{M} K_{mt}}$$

where

$\eta_m = v_m / \omega_m = \alpha_m (1-\alpha_i)/\alpha_i\ (1-\alpha_m)$, implying $\eta_m = 1$ when $\alpha_m = \alpha_i$;

$\lambda_{mt} = \lambda_{mt}^k / \lambda_{mt}^h$, implying λ_{mt} =1 when $\lambda_{mt}^k = \lambda_{mt}^h$.

Like in (5), we can rewrite (3) as:

$$(6) \qquad v_1 \lambda_{1t}^k \frac{Y_{1t}}{K_{1t}} = ... = \frac{Y_{it}}{K_{it}} = ... = v_M \lambda_{Mt}^k \frac{Y_{Mt}}{K_{Mt}} = \frac{\sum_{m=1}^{M} v_m \lambda_{mt}^k Y_{mt}}{\sum_{m=1}^{M} K_{mt}}$$

Combining (5) and (6) yields the following relationship between output and factor shares for reference member i of the integrated economy:

$$(7) \qquad \frac{Y_{it}}{\sum_{m=1}^{M} v_m \lambda_{mt}^k Y_{mt}} = \frac{K_{it}}{\sum_{m=1}^{M} K_{mt}} = \frac{H_{it}}{\sum_{m=1}^{M} \eta_m \lambda_{mt} H_{mt}} \qquad i = 1, \ldots, M$$

We term equation (7) the “equal-share” relationship. This relationship determines the distribution of output and the distribution of factors across M members of an integrated economy. Expression (7) contains both observable variables (Y_{mt}, K_{mt}, H_{mt}) and unknown parameters $(\alpha_m, \lambda_m^k, \lambda_m^h)$. Differences in technology or factor market imperfections imply a multiplicative rescaling of the observable variables that is different for each ratio. For example, a difference in α’s leaves the observed values (and share) of physical capital unaffected but transforms the observed values of output and human capital in different ways (through v_m and η_m respectively). If we assume that the M members of the integrated economy share the same technology ($\eta_m = \nu_m = \omega_m = 1$), and that there is costless (perfect) mobility of factors ($\lambda_{mt}^k = \lambda_{mt}^h = 1$), then we obtain the simplest expression of the equal-share relationship for any member i:

$$\text{(8)} \qquad \frac{Y_{it}}{\sum_{m=1}^{M} Y_{mt}} = \frac{K_{it}}{\sum_{m=1}^{M} K_{mt}} = \frac{H_{it}}{\sum_{m=1}^{M} H_{mt}} \qquad i = 1, \ldots, M$$

Hence, with perfect capital mobility and similar technology, each economy's share of total FIE output, and each economy's share of total FIE physical capital stock, equals its share of the total FIE stock of human capital.

Relationship (8) has an important implication. It contrasts the policies pursued in isolation by any given FIE member with those that are instead pursued jointly (harmonized) across members. For example, (8) remains unchanged when a coordinated educational policy by all FIE members increases their human capital by the same proportion. In contrast, the same policy implemented by only one member increases that member's share of total FIE human capital (as long as this policy is not imitated by other members). Hence, if FIE members have harmonized economic and social policies (e.g., fiscal, education, industrial policies) then the equal-share property implies that the relative performance of each member remains unaffected by these policies. In this sense, member shares can be considered to be a random variable whose outcome is dependent on the particular state of nature at time t. Such randomness can easily be understood from the fact that various kinds of random shocks, like discoveries, weather, or natural disasters, including some that are specific to a particular member, would give rise to new and different sets of shares for all members.

RANK-SHARE DISTRIBUTIONS AND ZIPF'S LAW

A rank-share distribution describes a particular relationship between the share and rank of a variable across a set of observational units. It is related to the concept of a rank-size distribution. For instance, a rank-size distribution for city size exists if the relationship between the natural logarithm of size and of rank is linear and exhibits a negative slope. Zipf's law arises when the slope value equals -1.

The existence of Zipf's law for city sizes is a widely documented empirical regularity.[4] Several explanations have been advanced for the observed regularity of Zipf's law with respect to the distribution of city sizes. Some argue it constitutes an optimal spatial pattern that arises when congestion and urbanization externalities interact as part of the process of development and growth of cities. Such forces are usually found in core models of urban and regional growth[5]. Others have stressed more mechanical forces that often involve a random growth process for city size. A recent example is Gabaix (1999), who draws on Gibrat's law[6] to assume that cities follow a random but common growth process. Normalizing city population by a country's total population, Gabaix shows (his Proposition 1) that if population shares evolve as geometric Brownian motion with an infinitesimal barrier then the steady state distribution of population shares will be a rank-size distribution that exhibits Zipf's law.

As previously noted, the equal-share property for members of an FIE, together with an assumed harmonization of FIE member's economic policies, implies that the relative performance of any one FIE member can be considered a random variable. Given this, we can adopt Gabaix's (1999) specification and assume that the share of variable j (e.g., j = output) evolves as geometric Brownian motion with a lower bound[7], and moreover, that the distribution of growth rates of these shares is common to all FIE

[4] See e.g. Brakman et al. (2001), Fujita et al. (1999), Gabaix and Ioannides (2004), Eeckhout (2004) and Rose (2005).

[5] For example, see Eaton and Eckstein (1997), Black and Henderson (1999), Brakman et al. (1999).

[6] Gibrat's law (Gibrat, 1931) states that firm growth is independent of firm size.

[7] One needs to prevent output and factors from falling below some lower bound in order to obtain a power law. Otherwise the distribution would be lognormal. A lower bound makes sense in integrated areas as important income transfers are institutionalized to prevent states/regions/countries to vanish. For example, the E.U. maintains a social fund and a regional fund.

members (i.e., Gibrat's law).[8] As in Gabaix (1999), this implies that the limiting distribution of the shares of variable *j* across FIE members will be a rank-share distribution that exhibits Zipf's law.

Empirical Specification

Consider a FIE consisting of M members. Let S_{mj} denote member m's share of the total FIE amount of variable j (j = output (y), physical capital (k) or human capital (h)) and let R_{mj} denote the rank of member m in the ranking of shares of variable j across all members ($m = 1,\ldots, M$). We assume $R_{mj} = 1$ for the member with the largest share of variable j and $R_{mj} = M$ for the member with the lowest share of variable j. If variable j has a rank-share distribution then we can write:

$$(9) \qquad S_{mj} = \gamma_j \left(R_{mj}\right)^{\beta_j}$$

where $\beta_j < 0$ is the power-law exponent and $0 < \gamma_j < 1$ is the share of variable j for the member with the highest rank (i.e., $R_{mj} = 1$). Zipf's law corresponds to $\beta_j = -1$, and it implies a specific relationship among member shares, namely: $S_{1j} = 2S_{2j} = 3S_{3j} = \ldots = MS_{Mj}$. This states, for example, that the share value of the highest ranked country is twice the share value of the second ranked country.

To empirically assess the hypothesis that output and factor shares conform to a rank-share distribution that exhibits Zipf's law we can take the natural logarithm of each side of (9) to obtain:

$$(10) \qquad \log\left(S_{mj}\right) = \theta_j + \beta_j \log\left(R_{mj}\right) + u_{mj} \qquad m = 1,\ldots,M;\ j = y,k,h$$

where $\theta_j = \log(\gamma_j) < 0$ and u_{mj} is an error term assumed to have the usual properties (i.e., i.i.d. with mean zero and constant variance). Estimates of the intercept and of the slope parameter in (10) are crucial to our analysis and are obtained by regressing the share of variable *j* on variable *j*'s rank value across FIE members.

---

[8] The equal-share relationship implies that the common expected rate of growth is zero since the sum over i of the output and factor shares in (8) must be 1.

We estimate (10) separately for the output share, physical capital share and human capital share with respect to the 51 U.S. states and the 14 E.U. countries. For U.S. states, we use annual cross-section data covering the period from 1990 to 2000. For E.U. countries the data instead consist of cross-sections equally spaced at 5-year intervals; these data generally cover the period from 1965 to 2000. The Appendix gives a complete description of the data.

Given estimates of (10) for a given dependent variable, evidence against Zipf's law can be assessed by testing if the estimated slope coefficient is significantly different from minus one. However, Gabaix and Ioannides (2004) and Nishiyama and Osada (2004) recently demonstrate that both the OLS estimate of β_j in (10) and its associated standard error are expected to be biased downward, with these biases diminishing as the number of observational units (M) increases. Hence, without some correction for these inherent biases, one is likely to more often reject Zipf's law when it is in fact true.

To correct for these biases, we follow Gabaix and Ioannides (2004, p. 10) and conduct, for the cases M = 14 (E.U. countries) and M = 51 (U.S. States), a Monte Carlo analysis of the OLS slope estimates derived from (10) under the assumption that Zipf's law holds.[9] The difference between the true slope value (-1) and the average of the OLS slope estimates gives an estimate of the downward bias, which is 0.172 for M = 14 and 0.081 for M = 51. Given these estimates of the bias for each M, an estimate of the true slope coefficient is obtained by adding the estimated bias to the OLS estimate derived from (10).

To obtain a bias adjusted estimate of the standard error we follow Nishiyama and Osada (2004) and use the asymptotic approximation to the true standard error of the OLS slope estimate given as $-\hat{\beta}_j \sqrt{2/M}$, where $\hat{\beta}_j$ is the OLS estimate of the slope in (10).[10]

[9] Briefly, for a given sample size M (either M = 14 or M = 51), 100,000 Monte Carlo simulations were performed drawing from an exact power law with coefficient 1 (Zipf's Law). This involved drawing M i.i.d. variables v_m, uniformly distributed in the interval [0, 1], and then constructing sizes $L_m = 1/v_m$. The L_m were then normalized into shares S_m that were then ordered and assigned a rank value R_m. We then preformed 100,000 OLS regressions using the specification $\log(S_m) = \theta + \beta \log(R_m) + u_i$. The complete results are available from the authors upon request.

[10] Another method for estimating the parameters of a power law distribution is the maximum likelihood Hill estimator (Hill, 1975). However, as Gabaix and Ioannides (2004) remark, the properties of the Hill estimator in finite samples can be "very worrisome," and in particular their theoretical results predict a large bias in parameter estimates and associated standard errors in small samples. We computed the Hill

The test statistic formed using these bias corrected values has asymptotically a normal distribution (Nishiyama and Osada, 2004).

Results

The first two columns of Table 1 report OLS estimates of (10) for the share of output, physical capital and human capital for the sample of U.S. states; the first two columns of Table 2 report the OLS estimates for the sample of E.U. countries.[11] Over both set of results, the adjusted *R*-squares fall in the range from 0.791 to 0.945, indicating a strong relationship between the share and rank of each variable.

Insert Tables 1 and 2 about here

In Table 1, the column labeled “Z-statistic Testing Slope = -1” indicates strong support for the hypotheses that the output and factor shares for U.S. states conform to a rank-share distribution that exhibits Zipf’s law; in no case can we reject (at the 5% level) the hypothesis that the slope coefficient is significantly different from –1. This is strong evidence that, for U.S. States, each of the three share distributions exhibit Zipf’s law.

Likewise, the column labeled “Z-statistic Testing Slope = -1” in Table 2 indicates also strong support for the hypotheses that the output and factor shares for E.U. countries conform to a rank-share distribution that exhibits Zipf’s law: we cannot reject (at 5% level) the hypothesis that the slope coefficient is significantly different from -1. These findings for U.S. states and for E.U. countries are striking empirical results.[12]

---

estimators (results not shown) and indeed found very high downward biases in both parameter estimates and standard errors.

[11] The standard errors associated with the OLS estimates are “robust” in the sense of White (1980).

[12] By comparison, we preformed similar tests for 30 developing countries and a “world” of 55 countries but no evidence of Zipf’s law was found at the usual significance levels.

FURTHER CHARACTERIZATION OF INTEGRATED ECONOMIES

The empirical findings of the preceding section have further implications regarding the characterization of integrated economies. One implication is the potential empirical validity of the equal-share relationship as derived in (8) since, if Zipf's Law holds, the output shares across countries, or shares of any given factor, are proportional. Hence, if the equal-share relationship holds for one country then it must then also hold for all other countries. A second implication is that if Zipf's Law holds then the limiting share values across FIE members are completely determined once the number of FIE members is specified.

Equal-share Relationship

A test for the equal-share relationship involves the null hypothesis given by equation (8) against the alternative hypothesis given by (7). Evidence in favor of the equal-share relationship can be obtained in two steps: (1) test for homogeneity of the OLS slope estimates (i.e., whether $\beta_y = \beta_k = \beta_h$) to verify that the distributions of shares come from a common power-law distribution and (2) test for homogeneity of the intercepts across the three share equations (i.e., whether $\theta_y = \theta_k = \theta_h$) to examine if the equal-share relationship holds with respect to the highest ranked member of each FIE (i.e., California for U.S. states and Germany for E.U. countries).[13] Failure to reject the null hypothesis would imply that technological differences and factor market imperfections are not strong enough to prevent the equal-share relationship from holding in a statistical sense.

Table 3 reports *p*-values for testing the hypotheses of slope homogeneity and of intercept homogeneity across the three share distributions in each sample year.[14] For U.S. states, in neither of the two years for which data are available on all three shares (1990 and 2000) can we reject the hypotheses of intercept equality and slope equality, supporting the equal-share relationship for U.S. states. The results for E.U. countries also

[13]Equally, it can be demonstrated that the equal-share property obtains if one assumes 1) that output shares alone exhibit Zipf's law and 2) that FIE members have identical, homogenous of degree one, production functions.

[14] These tests were performed by establishing, in each year, a system comprising the three share equations but without initially imposing any cross-equation parameter restrictions.

indicate support for the equal-share relationship. These test results are based on slope estimates uncorrected for bias. However, correcting for the expected downward bias would only strengthen the support for the equal-share relationship evidenced here.

Insert Table 3 About Here

Limiting Distribution of Shares

Let V_{mj} denote the level of variable j for member m. Assume, without loss of generality, that member 1 has the highest value of variable j and let δ_{mj} be member m's value of variable j relative to that of member 1 (i.e., $\delta_{mj} = V_{mj} / V_{1j}$), so that $\delta_{1j} = 1$. Now order the values of variable j in descending order. This ordering of the values of variable j across the $m = 1, \ldots, M$ members can be written:

$$(11) \qquad V_{1j} > \delta_{2j} V_{1j} > \delta_{3j} V_{1j} > \ldots > \delta_{Mj} V_{1j}$$

.

Since the total FIE amount of variable j is $(1 + \delta_{1j} + \delta_{2j} + \ldots + \delta_{Mj})V_{1j}$, (11) implies the following relations between member ranks and shares:

$$(12) \qquad \begin{aligned} &Rank\ 1: \ S_{1j} = \frac{1}{1+\delta_{2j}+\delta_{3j}+\ldots+\delta_{Mj}}; \\ &Rank\ 2: \ S_{2j} = \frac{\delta_{2j}}{1+\delta_{2j}+\delta_{3j}+\ldots+\delta_{Mj}}; \\ &Rank\ 3: \ S_{3j} = \frac{\delta_{3j}}{1+\delta_{2j}+\delta_{3j}+\ldots+\delta_{Mj}}; \\ &\qquad \vdots \qquad\qquad\qquad \vdots \\ &Rank\ \mathrm{M}: \ S_{Mj} = \frac{\delta_{Mj}}{1+\delta_{2j}+\delta_{3j}+\ldots+\delta_{Mj}}. \end{aligned}$$

Expressions (12) indicate that sequence of shares S_{mj} is a Harmonic series, where each share value S_{mj} depends on the values of the δ's and the number of members M. Accepting our preceding empirical evidence that the distribution of shares exhibits Zipf's law then $\delta_{2j} = 1/2$, $\delta_{3j} = 1/3$, $\delta_{4j} = 1/4$, etc., so that the theoretical shares in (12) can be computed once the number of members (M) is specified. For example, the theoretical share values for the M = 51 U.S. states are: 0.2213, 0.1106, 0.0738, 0.0553, …, 0.0043. For the M = 14 E.U. countries the theoretical share values are: 0.3075, 0.1538, 0.1025, 0.0769, …, 0.0220.

We conduct correlation and graphical analyses to gain insight on whether the observed distribution of shares conforms to the theoretically expected distribution of shares computed using (12). The relationship between actual shares and those computed from (12) is investigated in Table 4 which reports simple correlation coefficients between the natural logarithms of these shares for U.S. states and E.U countries in 1990 and 2000. The correlations range from 0.9176 to 0.9619 and all are highly significant, indicating a strong positive relationship between actual and theoretical shares.

Insert Table 4 about here

Figure 1 provides a graphical presentation of the share distributions by plotting the logarithm of the theoretically expected shares assuming Zipf's law holds and the logarithm of the actual shares in 2000 for each integrated economy. By definition, the theoretical shares (in logs) lie on a straight line with slope -1. Examination of the figures indicates that similar patterns between actual and theoretical shares appear for all three variables, whether for U.S. states or for E.U. countries. For example, for U.S. states, the graphs indicate that the share for the first observation (rank 1) is below the theoretical first share while in the middle range of the distribution the actual share is above the theoretical share. For E.U. countries the actual first share is instead very close to the theoretical share.

Insert Figure 1 About Here

There are several explanations for the observed deviation in actual share values from their theoretical values. One is that the theoretical share distribution is a steady state prediction and our sample values may not represent this ideal. Another is that our theoretical model assumes that each FIE is "closed," in that goods and factor flows arise only between FIE members. In reality, both U.S. states and E.U. countries have important trade and factor flow linkages with entities outside these defined integrated economies. A third is that, since the shares for a given variable sum to unity across observations, the sum of their differences at each rank (i.e., the "residual") must be zero. Hence, the sum of any positive "residuals" must be offset by the sum of negative "residuals." To an approximation, this same result will hold for the sum of the difference between the shares at each rank when measured in logarithms.

DISCUSSION

We examined empirically for evidence that the distribution of output and factor shares exhibit Zipf's law with respect to two integrated economies: the 51 U.S. states and 14 E.U. countries. The findings indicate that Zipf's law indeed holds for the distribution of these shares among U.S. states and also among E.U. countries. While there may be several explanations for this empirical finding, the evidence on the empirical significance of Zipf's law is consistent with a model that assumes that the growth process of the shares of members of an integrated economy is random and homogeneous across members.

Our empirical results also supported the existence of the equal-share relationship for both U.S. states and E.U. countries. This evidence leads to several implications regarding the characterization of integrated economies. First, the equal-share relationship addresses Lucas' (1990) question as to why capital does not flow from rich to poor countries. Namely, an economy with a low level (and hence a low share) of human capital will also have a low share of physical capital, and also a low share of output. Second, if the equal-share relationship holds, then all members of an integrated economy will have the same output per efficiency unit of labor. This implication is the essence of the absolute convergence hypothesis (Barro and Sala-i-Martin, 2004, p.47), here interpreted in terms of efficiency units of labor, not in per capita terms. Finally, the empirical significance of the equal-share relationship is consistent with the relative growth performance of members of an integrated economy being largely random, and hence strongly dependent on particular states of nature. Such randomness will be more true the greater the extent of economic integration among members, perhaps most exemplified by the integrated economy comprising U.S. states. Hence, it is more likely to be true the more harmonized are education systems and fiscal codes, when members do not run independent monetary policies, and when industrial policies are quickly imitated across members.

We also derived the result that, when Zipf's law holds, the values of the output and factor shares for members of a fully integrated economy are completely determined once the number of members is specified. These shares are limiting values that derive from the relative position (rank) of each member and would be expected to emerge as integrated economies approach full integration. Nonetheless, a comparison of actual share values to these theoretically expected share values indicated a high degree of agreement.

In providing evidence for Zipf's law and the equal-share relationship with respect to members of an integrated economy, this paper indicates that these empirical characterizations should be kept in mind when studying the implications of alternative policies on the relative growth of members of an integrated economic area.

REFERENCES

Baldwin, R. and P. Martin, 2004, "Agglomeration and Regional Growth," in Henderson, J.V. and Thisse, J.-F. (Eds.), *Handbook of Regional and Urban Economics-Volume IV* (Amsterdam: North-Holland), Chapter 2.

Barro, R. J., 1999, "Determinants of Democracy," *Journal of Political Economy* 107, S158-S183.

Barro, R. J. and J.W. Lee, 1993, "International Comparisons of Educational Attainment," *Journal of Monetary Economics* 32, 363-394.

________, 1996, "International Measures of Schooling Years and Schooling Quality," *American Economic Review* 86, 218-223.

________, 2000, "International Data on Educational Attainment: Updates and Implications," Center for International Development Working Paper 42, Harvard University.

Barro, R. J. and X. Sala-i-Martin, 2004, *Economic Growth* (Cambridge, MA: MIT Press).

Black, D. and V. Henderson, 1999, "A Theory of Urban Growth," *Journal of Political Economy* 107, 252-284.

Bowen, H. P., E. E. Leamer and L. Sveikauskas, 1987, "Multicountry Multifactor Tests of the Factor Abundance Theory," *American Economic Review*, 77 (5), 791-809.

Brakman, S., H. Garretsen and C. van Marrewijk, 2001, *An Introduction to Geographical Economies* (Cambridge: Cambridge University Press).

_______ and M. van den Berg (1999), The Return of Zipf: Towards a further understanding of the Rank-Size Rule, *Journal of Regional Science*, 39, 183-213.

Devereux, M.B. and B.J. Lapham, 1994, "The Stability of Economic Integration and Endogenous Growth," *Quarterly Journal of Economics* 109(1), 299-308.

Eaton, J. and Z. Eckstein, 1997, "Cities and Growth: Theory and Evidence from France and Japan," *Regional Science and Urban Economics* 27, 443-474.

Easterly, W. and R. Levine, 1998, "Africa's Growth Tragedy: Policies and Ethnic Divisions," *Quarterly Journal of Economics* 112, 1203-1250.

Eeckhout, J., 2004, "Gibrat's Law for (All) Cities", *American Economic Review*, 94(5), 1429-1451.

Fujita, M., P.R. Krugman and A.J. Venables, 1999, *The Spatial Economy: Cities, Regions, and International Trade* (Cambridge, Mass.: MIT Press).

Gabaix, X., 1999, "Zipf's Law for Cities: An Explanation," *Quarterly Journal of Economics* 114(4), 739-767.

_______ and Y. M. Ioannides, 2004, "The Evolution of City Size Distributions," in Henderson, *J*.V. and Thisse, *J*.-F. (Eds.), *Handbook of Regional and Urban Economics-Volume IV* (Amsterdam: North-Holland), Chapter 7.

Garofalo, G. and S. Yamarik, 2002, "Regional Convergence: Evidence from a New State-by-State Capital Stock Series", *Review of Economics and Statistics* 82, 316-323.

Gibrat, R., 1931, *Les Inegalites Economiques*, (Paris: Librairie du Recueil Sirey).

Grossman, G.M., and E. Helpman, 1991, *Innovation and Growth in the Global Economy* (Cambridge: MIT Press).

Hall, R.E. and C. I. Jones, 1999, "Why Do Some Countries Produce So Much More Output Per Worker than Others?," *Quarterly Journal of Economics* 114, 83-116.

Helpman, E. and P. Krugman (1985): *Market Structure and Foreign Trade*, (Cambridge MA: MIT Press).

Heston, A. and R. Summers, 1991a, "The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950-1988," *Quarterly Journal of Economics* 106, 327-368.

________, 1991b, *Penn World Table Version 5.6*, Center for International Comparisons at the University of Pennsylvania (CICUP).

________ and B. Aten, 2002, *Penn World Table Version 6.1*, Center for International Comparisons at the University of Pennsylvania (CICUP).

Hill, B. M., 1975, "A Simple Approach To Inference About The Tail Of A Distribution", *Annals of Statistics* 3:1163—1174.

Leamer, E. E., 1984, *Sources of International Comparative Advantage: Theory and Evidence*. Cambridge, MA : MIT Press.

Lucas, R. E., 1990, "Why Doesn't Capital Flow from Rich to Poor Countries?" *American Economic Review* 80, 92-96.

Mankiw, M.G., D. Romer and D. M. Weil, 1992, "A Contribution to the Empirics of Economic Growth," *Quarterly Journal of Economics* 107, 407-437.

Munnell, A., 1990, "Why Has Productivity Growth Declined? Productivity and Public Investment," *New England Economic Review,* 407-438.

Nishiyama, Y. and S. Osada, 2004, “Statistical Theory of Rank Size Rule Regression under Pareto Distribution,” Interfaces for Advanced Economic Analysis Discussion Paper, Kyoto University.

Rajan, R.G. and L. Zingales, 1998, “Financial Dependence and Growth,” *American Economic Review* 88, 559-586.

Ramey, G., and V. A. Ramey, 1995, “Cross-country Evidence on the Link between Volatility and Growth,” *American Economic Review* 85, 1138-1151.

Rivera-Batiz, L. and P. Romer, 1991, “Economic Integration and Endogenous Growth”, *Quarterly Journal of Economics* 106(2), 531-555.

Rose, A.K., 2005, “Cities and Countries,” NBER Working Paper 11762, (November).

Sachs, J. D. and A. Warner, 1995, “Economic Reform and the Process of Global Integration,” *Brookings Papers on Economics Activity,* 1-95.

Timmer, M. P., G. Ypma, and B. van Ark, 2003, “IT in the European Union: Driving Productivity Divergence?” *GGDC Research Memorandum* GD-67, University of Groningen.

U.S. Bureau of Economic Analysis, 2002, *Fixed Assets and Consumer Durable Goods for 1925-2001*.

Viaene, J.-M., and Zilcha, I., 2002, “Capital Markets Integration, Growth and Income Distribution,” *European Economic Review* 46, 301-327.

White, H., 1980, “A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity,” *Econometrica* 48, 817-838.

APPENDIX – DATA METHODS AND SOURCES

The output for each of the 51 U.S. states is measured by real gross state product as reported by the U.S. Bureau of Economic Analysis (BEA).[15] These data are available yearly from 1990 to 2000.

Estimates of state physical capital stocks are derived from BEA (2002) estimates of the total U.S. physical capital stock in each of nine one-digit industrial sectors comprising all economic activity.[16] These national physical capital stocks in each industry are allocated to each state by multiplying an industry's total capital stock[17] by that industry's contribution to a state's total income.[18] These industry capital stock estimates are then summed, for each state, to obtain an estimate of a state's total stock of physical capital.[19] The calculation performed for each state at each time t can be expressed algebraically as

$$k_m(t) = \sum_{j=1}^{9}\left[K_j(t)\left(y_{mj}(t)/Y_m(t)\right)\right]$$

In this equation, $k_m(t)$ is the stock of physical capital in state m, $y_{mj}(t)$ is value added by industry j in state m (m = 1…51), $Y_m(t)$ is state m's total value added, and $K_j(t)$ is the national level stock of physical capital in industry j (j = 1,…, 9). This procedure assumes that the capital-to-output ratio within an industry j (i.e., $k_{mj}(t)/y_{mj}(t)$) is the same across U.S. states, that is, $k_{mj}(t)/y_{mj}(t) = K_m(t)/Y_m(t)$. In turn, this assumption implies that an industry is in a common steady state across all U.S. states.[20] For example, the agricultural sector in Texas is in the same steady state as its counterpart in Oregon, and

[15] Data on gross state product available at http://www.bea.doc.gov/bea/regional/gsp

[16] The sectors (BEA code) are Farming (81), Agricultural services, forestry, fishing & other (100); Mining (200); Construction (300); Manufacturing (400); Transportation(500); Wholesale and retail trade (610); Finance, insurance and real estate (700); and Services (800).

[17] Data on state physical capital stocks by industry were taken from U.S. Fixed Assets Tables, available at http://www.bea.doc.gov/bea/dn/faweb

[18] Annual data on state value added available at http://www.bea.doc.gov/bea/regional/spi

[19] This procedure follows that used by Munnel (1990) and Garofalo and Yamarik (2002).

[20] If a sector is converging towards its steady state, the output-to-capital ratio would be below its steady-state value. This only poses a problem if the initial output-to-capital ratios vary across U.S. states. If the ratios do vary, the procedure would allocate too much to those states further from steady-state and too little to those states closer to their steady state.

the manufacturing sector in Pennsylvania is in the same steady state as its counterpart in Ohio.[21] The constructed physical capital data are from 1990 to 2000, on a yearly basis.

Human capital stocks for U.S. states are proxied by the number of persons with at least secondary level education. They are derived from data on state educational attainment taken from the U.S. Bureau of the Census.[22] Census data on educational attainment are available only every 10 years, which limited the construction of human capital stocks to two years: 1990, and 2000.

For E.U. countries, each country's total output is measured by its real gross domestic product (GDP) derived from the data on real GDP per capita (base year = 1996) and population in Penn World Tables 6.1 (Heston, Summers and Aten, 2002). The output data are available annually from 1960 to 2000.

Data on E.U. physical capital stocks are derived from Penn World Tables 5.6 (Heston and Summers, 1991a and 1991b) which reports four data series for each country: (1) population, (2) physical capital stock per worker, (3) real GDP per capita and (4) real GDP per worker. The physical capital stocks for each country are constructed as the product of the first three series divided by the last series. The data covers the period 1965-1990. Physical capital stock data for E.U. countries are also available from Timmer et al. (2003) covering period 1980-2000.[23] These data sources are combined to have physical capital stock data in each of seven years from 1965 to 2000.[24]

Each E.U. country's stock of human capital stock is measured by multiplying the percentage of a country's population having at least a secondary level of education with the country's total population. Data on the rate of educational attainment for each country are taken from Barro and Lee (1993, 1996, and 2000).[25] Data on a country's population are from Heston, Summers and Aten (2002). Since data on rates of educational attainment

---

[21] If a sector has a different steady state, and hence a different capital-to-output ratio, the procedure will allocate too much to states with lower ratios and too little to states with higher ratios. However, this possibility is unlikely if competition lead firms in all states to adopt the best available production technology.

[22] Decennial Census dataset available at http://factfinder.census.gov

[23] Physical capital database available at http://www.ggdc.net/dseries/growth-accounting.shtml

[24] Estimation was conducted using both sets of data for E.U. countries. No qualitative difference in results was found for the years in which data were available from both sources (i.e., 1980, 1985 and 1990). For these three years we therefore report only the results using the capital stock data from Timmer et al. (2003).

[25] Other studies using the Barro-Lee data include Rajan and Zingales (1998), Ramey and Ramey (1995), Barro (1999), Easterly and Levine (1998), Hall and Jones (1999) and Sachs and Warner (1995).

are only available every 5 years, the data sample is limited to five-year intervals from 1960 to 2000. Following this constraint, the output and physical capital stocks are also obtained in five-year intervals.

The 14 E.U. countries are: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Portugal, Spain, Sweden and United Kingdom.[26]

---

[26] Luxembourg is excluded for lack of data on human capital. Given the small scale of Luxembourg's economy relative to other E.U. countries this omission is unlikely to affect the E.U. results.

TABLE 1

OLS Estimates of Rank-Share Relationships for U.S. States

Variable	Year	Intercept [a]	Slope [b]	Z-statistic Testing Slope = -1 [c]	Adj. R^2
Output Share (M = 51)	1990	-1.179 (0.248)	-1.101 (0.081)	-0.092	0.887
	1991	-1.194 (0.248)	-1.093 (0.081)	-0.055	0.884
	1992	-1.199 (0.252)	-1.090 (0.082)	-0.042	0.883
	1993	-1.207 (0.258)	-1.085 (0.084)	-0.019	0.881
	1994	-1.208 (0.265)	-1.084 (0.086)	-0.014	0.876
	1995	-1.209 (0.265)	-1.083 (0.086)	-0.009	0.874
	1996	-1.205 (0.267)	-1.085 (0.087)	-0.019	0.872
	1997	-1.192 (0.271)	-1.091 (0.088)	-0.046	0.868
	1998	-1.173 (0.272)	-1.100 (0.088)	-0.087	0.868
	1999	-1.168 (0.271)	-1.103 (0.088)	-0.101	0.866
	2000	-1.164 (0.266)	-1.106 (0.087)	-0.114	0.868
Physical Capital Share (M = 51)	1990	-1.199 (0.246)	-1.092 (0.080)	-0.051	0.892
	1991	-1.207 (0.247)	-1.089 (0.080)	-0.037	0.891
	1992	-1.200 (0.251)	-1.092 (0.081)	-0.051	0.892
	1993	-1.197 (0.257)	-1.093 (0.083)	-0.055	0.890
	1994	-1.196 (0.266)	-1.092 (0.086)	-0.051	0.884
	1995	-1.173 (0.275)	-1.102 (0.089)	-0.096	0.879
	1996	-1.168 (0.276)	-1.105 (0.089)	-0.110	0.878
	1997	-1.126 (0.286)	-1.125 (0.093)	-0.198	0.870
	1998	-1.126 (0.283)	-1.126 (0.091)	-0.202	0.876
	1999	-1.108 (0.283)	-1.135 (0.092)	-0.240	0.875
	2000	-1.093 (0.282)	-1.143 (0.091)	-0.274	0.880
Human Capital Share (M = 51)	1990	-1.244 (0.280)	-1.064 (0.091)	0.081	0.854
	2000	-1.264 (0.293)	-1.054 (0.096)	0.129	0.839

[a] OLS standard errors in parentheses. All intercept coefficients significantly different from zero at 1% level.
[b] OLS standard errors in parentheses. All slope coefficients significantly different from zero at 1% level.
[c] Computed as the OLS slope estimate minus (- 1) plus 0.081 (the bias) divided by the asymptotic approximation of the true standard error (given as minus the OLS slope estimate times 0.198 = $(2/51)^{0.5}$. All slope coefficients are not significantly different from –1 at the 5% level.

TABLE 2: OLS Estimates of Rank-Share Relationships for E.U. Countries

Variable	Year	Intercept [a]	Slope [b]	Z-statistic Testing Slope = -1 [c]	Adj. R^2
Output Share (*M* = 14)	1960	-0.645 (0.397)	-1.461 (0.192)	-0.523	0.908
	1965	-0.665 (0.416)	-1.435 (0.204)	-0.485	0.889
	1970	-0.699 (0.433)	-1.406 (0.212)	-0.440	0.867
	1975	-0.742 (0.435)	-1.366 (0.211)	-0.376	0.859
	1980	-0.755 (0.419)	-1.357 (0.202)	-0.361	0.870
	1985	-0.763 (0.417)	-1.354 (0.199)	-0.356	0.872
	1990	-0.772 (0.420)	-1.346 (0.198)	-0.342	0.872
	1995	-0.777 (0.405)	-1.343 (0.187)	-0.337	0.878
	2000	-0.857 (0.376)*	-1.272 (0.170)	-0.208	0.885
Physical Capital Share (*M* = 14)	1965	-0.816 (0.417)	-1.293 (0.217)	-0.248	0.851
	1970	-0.825 (0.396)	-1.275 (0.208)	-0.214	0.858
	1975	-0.836 (0.388)*	-1.262 (0.203)	-0.189	0.858
	1980	-0.760 (0.484)	-1.332 (0.245)	-0.318	0.828
	1985	-0.732 (0.404) *	-1.358 (0.205)	-0.362	0.870
	1990	-0.670 (0.398)	-1.418 (0.206)	-0.459	0.873
	1995	-0.632 (0.330)	-1.457 (0.174)	-0.518	0.908
	2000	-0.658 (0.382)	-1.431 (0.186)	-0.479	0.904
Human Capital Share (*M* = 14)	1960	-0.147 (0.448)	-2.103 (0.287)	-1.171	0.791
	1965	-0.343 (0.341)	-1.890 (0.184)	-1.005	0.880
	1970	-0.529 (0.280) *	-1.639 (0.176)	-0.754	0.865
	1975	-0.642 (0.236) **	-1.518 (0.126)	-0.603	0.928
	1980	-0.683 (0.239) **	-1.433 (0.122)	-0.482	0.933
	1985	-0.747 (0.185) **	-1.409 (0.092)	-0.445	0.945
	1990	-0.895 (0.191) **	-1.241 (0.112)	-0.147	0.912
	1995	-0.897 (0.201) **	-1.225 (0.115)	-0.114	0.912
	2000	-0.905 (0.196) **	-1.215 (0.110)	-0.094	0.919

[a] OLS standard errors in parentheses. Significantly different from zero at ** = $p < 0.05$ or * = $p < 0.10$
[b] OLS standard errors in parentheses. All slope coefficients significantly different from zero at the 1% level.
[c] Computed as the OLS slope estimate minus (-1) plus 0.172 (the bias) divided by the asymptotic approximation of the true standard error (given as minus the OLS slope estimate times 0.3779 = $(2/14)^{0.5}$. All slope coefficients are not significantly different from –1 at the 5% level.

TABLE 3

Results Testing the Equal-Share Relationship

Integrated Economy	Year	*p*-values for testing across-equation homogeneity of	
		intercepts	slopes
U.S. States	1990	0.9680	0.9014
	2000	0.8241	0.5964
European Union	1965	0.6063	0.0445 [a]
	1970	0.8011	0.2797
	1975	0.8619	0.3655
	1980	0.9689	0.8461
	1985	0.9969	0.9305
	1990	0.8111	0.6034
	1995	0.7124	0.3697
	2000	0.7291	0.4072

[a] Cross-equation homogeneity rejected at 5% level.

TABLE 4

Correlations between Logarithm of Actual and Theoretical Output and Factor Shares for U.S. States and E.U. Countries, 1990 and 2000

Integrated Economy	Year	Correlation Between Logarithm of Actual Shares and Theoretical Shares of		
		Output	Physical Capital	Human Capital
U.S. States	1990	0.9429	0.9456	0.9258
	2000	0.9332	0.9393	0.9176
European Union	1990	0.9392	0.9397	0.9397
	2000	0.9453	0.9548	0.9619

FIGURE 1: Theoretical and Actual Share Distributions for U.S. States and E.U. Countries

U.S. States	E.U. Countries